Abstract

HMG1 and 2 (high mobility group proteins 1 and 2; renamed HMGBl and 2) contain two DNA-binding HMG-box domains (A and B) and a long acidic C-terminal domain. They bind DNA without sequence specificity, but have a high affinity for bent or distorted DNA, and bend linear DNA. The individual A and B boxes (which, although broadly similar, show both structural and functional differences) exhibit many of the structure-specific properties of the whole protein. The acidic tail modulates the affinity of the tandem HMG boxes in HMGl and 2 for a variety of DNA targets, including four-way junctions, but not distorted DNA minicircles, to which the proteins bind with very high affinity. HMGl and 2 appear to play important architectural roles in the assembly of nucleoprotein complexes in a variety of biological processes, for example V(D)J recombination, the initiation of transcription, and DNA repair.

Key words: DNA distortion, DNA minicircles, high mobility group (HMG) proteins, HMG box.

Introduction

HMG1 and 2 (high mobility group proteins 1 and 2; recently redesignated HMGBl and 2 [1]) are ubiquitous proteins in vertebrates. They are relatively abundant (about 1 molecule per 10–15 nucleosomes on average) and, like histones, bind to DNA without sequence specificity, being initially regarded as probable chromatin structural components [2]. However, more recent studies have revealed other important roles for HMGl and 2, which result from their distinctive ‘structure-specific’, rather than sequence-specific, DNA-binding properties. They bend linear DNA and bind preferentially to bent or distorted DNA in vitro, and appear to have a role in vivo in the assembly of nucleoprotein complexes [2–4].

HMG1 and 2 have two distinguishing features: two HMG boxes (A and B), homologous folded domains of about 80 amino acid residues which mediate DNA binding, and a long acidic tail containing 30 (HMG1) or 20–24 (HMG2a and 2b) aspartic or glutamic acid residues, linked to the boxes by an overall basic region of about 20 residues (Figure 1). Abundant HMG-box proteins in organisms other than vertebrates (e.g. HMG-D and HMG-Z in Drosophila melanogaster and Nhp6ap and Nhp6bp in Saccharomyces cerevisiae),
assumed to be at least to some extent the functional counterparts of HMG1 and 2, have only a single HMG box which may (as in HMG-D and -Z; 12-residue tail) or may not (as in Nhp6ap and Nhp6bp) be accompanied by an acidic tail [4]. However, both *Drosophila* and yeast also contain less abundant tandem-HMG-box proteins (DSP1, and Hmolp and Hmol2p, respectively), which may have specialized roles. Characterized single HMG-box domain proteins from plants contain basic extensions, N-terminal and sometimes C-terminal to the HMG box, and (often relatively short) C-terminal acidic tails [5].

**Figure 1**
The domain structure of HMG1, HMG2a and HMG2b (from chicken erythrocytes)

A and B denote the HMG boxes and C the acidic C-terminal domain ("tail"). The connecting lines denote basic linkers. In HMG2a both are shorter by two residues than the linkers in HMG2b. Numbers refer to amino acid residues.

Role of the HMG-box domains

The unusual DNA-binding properties of HMG1 and 2 *in vitro* are well characterized, and are the boxes, in single copies and not accompanied by acidic tails, are also found in a number of sequence-specific transcription factors such as SRY (the product of the sex-determining gene on the Y chromosome) and LEF-1 (lymphocyte enhancer factor 1). For recent reviews, see [3,4].

The structures of several HMG boxes from both sequence-specific and non-sequence-specific proteins have now been determined, some in complex with DNA (reviewed in [4,6]). They have a common global fold consisting of an 'L-shaped' structure with three α-helices, the long arm comprising the N-terminal extended strand and helix III, and the short arm comprising helices I and II. However, there are subtle structural differences between the A and B domain HMG boxes (Figures 2a and 2b), as well as functional differences *in vitro* (see below). The HMG boxes from SRY [7], LEF-1 [8], HMG-D [9] and Nhp6ap [10] resemble the B domain more closely than the A domain [11], which appears to occur uniquely in the canonical HMG-1 and -2 proteins.

**Figure 2**

(a) and (b) NMR structures of the A and B domains of HMG1 respectively [11,45], showing the residues at the two positions on the concave face known to be involved in intercalation into the minor groove of DNA in the structures of other HMG boxes (see text). Note that whereas both Phe19 (primary position) and Ile34 (secondary position) would probably intercalate in a B-box–DNA complex (although there is no structure yet), in the A domain only the Phe at position 37 would be competent to intercalate. Note also the structural differences between the A and B domain, especially in the orientation of helices I and II and the intervening loop region. (c) Structure of the LEF-1 HMG box–DNA complex determined by NMR spectroscopy [8] (drawn from coordinates in the PDB). In all cases the protein is displayed using MOLSCRIPT [46].

© 2001 Biochemical Society
key to their biological roles. Like all HMG-box proteins, HMG1 and 2 bind DNA and bind preferentially to distorted DNAs such as four-way junctions, kinked cisplatin-modified DNA and DNA 'bulges' [6]. They also bind to cross-overs in supercoiled DNA and can cause looping of linear DNA. Many of these features are also properties of single HMG-box domains and thus originate in the HMG boxes themselves.

Because of the difficulties in forming specific protein–DNA complexes with non-sequence-specific HMG-box domains, detailed insights into the nature of the binding and mechanism of DNA distortion came first from studies of complexes containing the sequence-specific HMG boxes of SRY [7,12] and LEF-1 [8]. The HMG box binds through its concave face (the region encompassed by the arms of the L) to the minor groove of B-form DNA (Figure 2c), and causes bending by partial intercalation of a hydrophobic residue close to the N-terminus of helix I, on the concave face of the HMG box (Ile in SRY and Met in LEF-1) between base pairs. This opens up the minor groove, thus increasing the protein–DNA interface, and kinks the DNA towards the opposing major groove which is thereby compressed (Figure 2c). More recently structures or detailed models for non-sequence-specific HMG box–DNA complexes have become available [9,10,13–15], which in general reiterate the main features of interaction but also reveal interesting and significant differences.

The structure of the non-sequence-specific HMG box of HMG-D in complex with an 11 bp oligonucleotide revealed an intercalating residue (Met) in the position corresponding to the Ile in SRY and Met in LEF-1 (the ‘primary’ intercalating position), but unexpectedly revealed an additional intercalation by a hydrophobic residue (Val), and the adjacent Thr, immediately before the beginning of helix II (the ‘secondary’ intercalating site) [13,15]. A similar interaction involving a Phe residue was also postulated. Similar features, with two intercalating sites, were evident in an NMR-based model of the HMG box of yeast Nhp6ap, like HMG-D a ‘B-type’ HMG box (see above), bound to a 15 bp oligonucleotide [10]. There is no structure of the B domain of HMG1 itself bound to DNA, but it has Phe and Ile residues at the primary and secondary positions respectively (Figure 2b), and is thus potentially able to intercalate in the same way as the HMG boxes of HMG-D and Nhp6ap. The A domain of HMG1, however, is unique: it has Ala at the primary position rather than a bulky hydrophobic residue capable of intercalation (Figure 2a). However, there is such a residue (Phe37) at the putative secondary intercalating position in the A domain. Although there is no structure of the A domain bound to linear DNA, a crystal structure of the domain bound to an oligonucleotide kinked at a cis-platinum adduct [14] revealed a special role for Phe37, which partially intercalates within the kinked platinated GpG site. Bending of linear DNA by the A domain would thus be expected to involve intercalation by Phe37.

The proposed differences in intercalation between the A and B domains of HMG1 would explain why the B domain bends linear DNA better than the A domain [16,17]. However, the A domain binds more tightly than the B domain to four-way junction DNA and resists competition by a large excess of supercoiled DNA, which is itself a preferred substrate [17]. The A domain binds preferentially to four-way junctions even when the B domain is present, as in the AB didomain [18]. The high affinity for the junction is due partly to the electrostatic complementarity between the positively charged end of the short arm of the A-domain HMG box and the ‘hole’ at the centre of the junction, and partly to the proposed stacking of Phe37 against a base at the proximal end of one of the junction arms.

Many HMG-box proteins have a basic extension either C- or N-terminal to the HMG box, which increases the DNA-binding affinity of the HMG box. In the protein–DNA complex containing the LEF-1 HMG box (Figure 2c), as well as those containing the Nhp6ap and HMG-D boxes, this region binds in the compressed major groove on the face of the helix opposite to the widened minor groove [8,10,19]. This stabilizes the bend and facilitates circularization [20,21].

**The acidic tail**

The cellular roles of the distinctive long acidic tails of HMG1 and 2 are not clear. However, in vitro they lower the affinity of the HMG boxes for most DNA substrates, such as linear DNA, supercoiled DNA and four-way junctions [22–24] (although not DNA minicircles; see below). Whether this is a consequence of interaction of the tail with one or both of the HMG boxes [25], or simply charge repulsion, is not clear. The length of the tail is an important factor in vitro, giving HMG2 variants HMG2a and 2b (which have tails of 20–24 acidic residues) slightly different properties from
Figure 3

An architectural and chaperone role for HMG1 in facilitation of transcription factor binding

A similar mechanism could apply to the role of HMG1 in V(D)J recombination (see text). The major features are: (a) Recruitment of HMG1 (shown hatched) through interaction with a sequence-specific protein (TF1) (protein-protein interactions have been detected in several cases). (b) Formation of a ternary complex at the TF1 target site. Such complexes are likely to be unstable but have been detected/inferred in some cases; in the ternary complex HMG1 may bend the DNA, thus providing the potential for the recruitment of additional DNA-binding proteins to the complex. (c) If no additional protein is recruited, HMG1 may dissociate from the complex leaving the transcription factor stably bound: in this schematic, the final DNA is shown straight (as in the case of Oct-1 binding), but this is not necessarily the case (e.g. the progesterone receptor bends its DNA in the binary complex). (d) Alternatively, a second protein may be recruited, followed optionally (e) by HMG1 dissociation. (Note that only a single HMG box is shown, since the separate A and B HMG boxes of HMG1 have been shown to facilitate binding of various transcription factors in vitro. However, in other cases [e.g. V(D)J recombination], both HMG boxes might be required. In this case, one box may be involved in protein-protein interactions, the other in bending the DNA; see the text for further details. Adapted from Trends. Biochem. Sci. 26, J. O. Thomas and A. A. Travers, "HMG1 and 2, and related ‘architectural’ DNA-binding proteins", pp. 167–174, © 2001, with permission from Elsevier Science.)
HMG1 (30 consecutive Asp or Glu residues). In a study of HMG1, HMG2a and HMG2b from chicken erythrocytes (to be known in future as HMGB1, HMGB2 and HMGB3; [1]), HMG1 was found to be less effective than HMG2a and 2b (at a given molar input ratio) in supercoiling relaxed, closed-circular DNA; in inducing ligase-mediated circularization of an 88 bp DNA fragment that would not circularize without protein-induced bending; and in binding to four-way DNA junctions [24]. Removal of the acidic tail largely abolishes the differences between the three proteins and increases the affinity of the HMG boxes for DNA. Replacement of the acidic tail of HMG1 with the tail of HMG2a or 2b gives hybrid proteins with essentially the DNA-binding properties of HMG2a or HMG2b respectively. The length (and possibly sequence) of the acidic tail thus appears to be the dominant factor in mediating the differences in DNA-binding properties in vitro between HMG1, 2a and 2b.

DNA minicircles formed in the presence of HMG-box proteins are highly distorted. HMG1 and 2 bind to 88 bp circles produced in this way with high affinity. They resist competition with up to a 20000-fold (by mass) excess of four-way junction DNA, or two-base bulged DNA, and even supercoiled DNA at 5000-fold excess (by mass) is only partially effective [24]. Surprisingly, removal of the acidic tail has little or no effect on the affinity of HMG1 and 2 for the circles; the affinity of the boxes for DNA minicircles must be far higher than any modulating interactions of the acidic tail with the HMG boxes. The tail is essential for selective binding to minicircles in the presence of roughly equimolar linear DNA (the affinity for which is reduced by the acidic tail) [24], as also shown for HMG-D [26].

Binding of HMG1 and 2 to the 88 bp minicircles is highly co-operative, with two molecules bound, co-operativity probably being mediated through the DNA [27]. A study of the binding of single and tandem boxes, both with and without basic extensions, revealed that two HMG boxes and the adjoining basic region (that links the HMG boxes to the acidic tail) are needed for strong co-operative binding limited two protein molecules [27]. The acidic tail makes little difference. DNA minicircles are probably not physiological targets for HMG-box proteins, but studies of circles and other DNAs are nonetheless likely to be relevant, given that the cell contains both a range of potential distorted DNA targets and a variety of HMG-box proteins with one or two boxes, and with or without basic extensions to the box(es) and/or an acidic tail.

The acidic domain of HMG1 and 2 may have functions in vivo in addition to modulation of binding (to all DNAs except minicircles) by the HMG boxes. It appears to be involved in the nuclear retention of HMG2 [28] and in the stimulation of transcription by HMG1 in cultured cells [29]; and the tail of HMG2 can replace the acidic ‘activation domain’ of HMG-14 in stimulation of transcription from simian virus 40 (SV40) minichromosomes and decondensation of chromatin structure in vitro [30]. The long tail of about 50 acidic residues in the RNA polymerase I transcription factor UBF1, which contains six HMG-boxes, has similar properties, being essential for transactivation [31,32] and nucleolar accumulation [33].

HMG1 and 2 as ‘architectural transcription factors’ and ‘chaperones’

The DNA-binding properties of HMG1 and 2 appear to make them useful components in various DNA transactions involving other proteins. Since the identification of HMG2 as a cellular component that facilitated binding of the ubiquitous transcription factor Oct-1 to its target DNA site [34], it has become clear that HMG1 and 2 can function similarly in vitro and in transfection assays, with a number of other sequence-specific transcription factors. These include HoxD9, p53, Rel proteins and steroid hormone receptors (see [4] for a summary and references). HMG1 and 2, which have no sequence specificity, may be recruited by protein–protein interactions to specific sites, either before (Figure 3a) or after the sequence-specific protein (TF1) interacts with its target DNA; however, indirect recruitment of HMG1 and 2, for example, to a bend generated by binding of the sequence-specific protein, might also occur. Evidence for ternary complexes (Figure 3b) that may exist only transiently is scant, and HMG1 might rapidly dissociate from such a complex, leaving TF1 stably bound to its target site (Figure 3c). Impaired binding of the (hormone-bound) glucocorticoid receptor, a transcription factor, would provide an explanation for the pleiotropic effects on glucose metabolism resulting from deletion of the mouse Hmg1 gene [35]. That the deletion was not lethal was probably due to some functional redundancy between members of the HMG1 and 2 family.
The precise role of the recruited HMG1 is unclear. One possibility is that it enhances DNA and improves contact between the transcription factor and DNA (which, as in the case of Oct-1/DNA, is not necessarily bent in the final complex). DNA bending might be particularly important if there were a second protein component (TF2) in vivo that interacted with the first, a scenario which a purified in vitro system would not address. A quaternary complex (Figure 3d), in which HMG1 plays a clear architectural role in facilitating the interaction of TF1 and TF2 bound to their respective target sequences, might be stable or might bind with higher affinity, as it does to four-spacers (the general electrostatic interactions [21]. The B domain is much less effective. However, if the other primarily in, for example, protein-protein interactions (although possibly also stabilizing the binding of the first domain through protein-protein contacts via the RAG1 homeodomain. Together with RAG1 and 2, they appear to bend the DNA between the two recombination signal sequences spaced by 23 bp and stabilize a nucleoprotein complex that ensures that V(D)J recombination occurs only between recombination signals with 12 and 23 bp spacers (the ‘12/23 rule’). Although individual HMG boxes are sufficient to promote binding of several transcription factors in vitro, in V(D)J recombination both HMG boxes might be needed [37].

In general, either the A or the B domain might, in principle, be involved primarily in DNA binding, depending on the DNA substrate, and the other primarily in, for example, protein-protein interactions (although possibly also stabilizing the binding of the first domain through general electrostatic interactions [21]. The B domain might be used for bending linear DNA, since the A domain is much less effective. However, if recognition of pre-distorted DNA is required, either A or B would be competent and, based on in vitro binding studies, in some cases the A domain might bind with higher affinity, as it does to four-way junctions.

**HMG1 and 2 in chromatin**

Facilitation of transcription factor binding, and other HMG1-mediated DNA transactions in eukaryotes, take place in a chromatin context, but probably often at nucleosome-free gaps, such as at promoters and enhancers, characterized by their DNase I hypersensitivity. But do HMG1 and 2 have any role in chromatin structure; for example, in bending DNA and/or stabilizing kinks? Although the DNA in the nucleosome core particle is less tightly curved than in HMG domain–DNA complexes, some models of chromatin folding require bending of the linker DNA. The DNA entering and exiting from the nucleosome, at sites flanking the dyad axis, might be sharply bent and provide possible HMG1 binding sites. This is also the general location of H1 in the nucleosome. The globular domain of H1 is located to one side of the dyad, bridging a point about 5 bp away from it and another close to the end of the 166 bp of chromosomal DNA [38], a manifestation of two DNA-binding sites on the globular domain. Through very approximate co-localization, HMG1 and histone H1 might thus have similar functional roles in some circumstances (for example, in chromosome condensation, when only one of the two is present at different developmental stages [39,40]), although they would recognize distinct architectural features in the nucleosome: H1 binds to a pair of duplexes and HMG1 to a putative kink.

**Concluding remarks**

A major role for the relatively abundant HMG1 and 2 class of proteins is to facilitate the formation of complex nucleoprotein assemblies, in much the same way that the sequence-specific HMG-box protein LEF-1 does at the T-cell receptor locus, TCRα [41]. The DNA in the assemblies may be tightly bent, and the role of the HMG proteins is to overcome the axial rigidity of the DNA [42]. However, because HMG1 and 2 bind to DNA without sequence specificity, they have to be actively recruited to particular sites, possibly by direct interaction with a sequence-specific protein, although indirect DNA-mediated effects are also possible. Natural aberrations in the cell to which HMG1 and 2 might bind include UV-induced pyrimidine dimers or sites of DNA repair, where they may interact with ligases and other enzymes (see [4] for references). The activities of HMG1 and 2 in all these processes could potentially be regulated by acetylation [43,44], but this has yet to be fully explored. Binding sites for HMG1 and 2 in chromatin generally are likely to be bends and kinks in the DNA at the exit and entry points to the nucleosome, or arising from folding into a higher-order structure.
References


Received 12 April 2001